



HAL
open science

From syntactic and semantic analysis to prosodic perspectives: an automatic approach

Sébastien Monnier, Anne Lacheret-Dujour, Michel Morel

► **To cite this version:**

Sébastien Monnier, Anne Lacheret-Dujour, Michel Morel. From syntactic and semantic analysis to prosodic perspectives: an automatic approach. 15th International Congress of Phonetic Sciences (ICPhS-15), Aug 2003, Barcelone, Spain. pp.1137-1140. hal-00012283

HAL Id: hal-00012283

<https://hal.science/hal-00012283>

Submitted on 19 Oct 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From syntactic and semantic analysis to prosodic perspectives: an automatic approach.

MONNIER Sébastien, LACHERET Anne and MOREL Michel

CRISCO¹, University of Caen

E-mail: monnier@crisco.unicaen.fr, morel@crisco.unicaen.fr, anne@lacheret.com

ABSTRACT

An absolute prosodic modelisation for an English text-to-speech synthesis system implies the study of three main levels: pragmatics, semantics and syntax. Such levels can provide accentuation potentials. Indeed, they show some possibilities but they cannot predict how a given sentence will be uttered. The purpose of this article is to demonstrate the main characteristics that are necessary to the prosodic marking of a text. Furthermore, we will show how this marking can improve the text-to-speech synthesis Kali, developed at the university of Caen.

Firstly, we will delimit the four prosodic levels we used in our study. Secondly, we will show how to locate these levels with a few textual clues including syntactic, semantic and pragmatic information. Thirdly, we will present the use of the linguistic development software Unitex and how it is complementary with Kali.

1. INTRODUCTION

Nowadays, syntactic difficulties are resolved when an intonational model for automatic generation of prosody is developed in a text-to-speech synthesis system. However, this cannot be applied to semantic and pragmatic modules. By semantic modules, we mean the description given by the lexicon for every feature needed: human / non-human, subjectivity... As for pragmatic aspects, they are linked to informational structures and to the attitudes the speaker wants to communicate.

It is difficult to model these elements without an efficient implement. The Unitex software, a linguistic development environment, makes this type of process possible. In this way, the user can indeed build lexicons and grammars that are described in a graphic form. They can thus be adapted to prosodic analysis. Once the graphs are assembled in systems, they limit and mark out particular structures.

Grammars can delimit syntactic groups in such a way as to mark boundary tones. Some specific structures, such as interpolated clauses can be analysed once the corresponding graphs are elaborated. As for the lexicons, they make it possible to highlight intensity markers

(*extremely, so much...*), words which convey expressivity and that are linked to attitudinal functions (*unbelievable, incredible...*) or words which are the most likely to carry new information (proper nouns, neologisms, technical words...).

In order to develop an intonosyntactic model for an English text-to-speech system, a complete description of these accentuation contexts is a must. For this purpose, we have created a semantic network which is completely directed towards prosodic analyses and enables us to take into account informational structure.

The purpose of this paper is first to present our theoretical background regarding semantic and pragmatic projections upon intonational structure, and to formalise a set of corresponding rules. Second, to show the way the software used operates to fulfil the function of implementing these rules and their integration in the Kali text-to-speech synthesis system developed at Caen University.

2. THEORETICAL APPROACH

Before setting out our analysis methods and the corresponding interpretations, we will explain our approach for the different levels of English intonative structures.

2.1. The stress

In French, word accentuation is fixed and is aligned on boundary accentuation. On the contrary, in English, accent location is determined by morphophonological rules. It is often on the penultimate or the antepenultimate syllable. This accent, called "stress", is the main level for the understanding of a discourse. As it depends on morphological rules, we can integrate it into a lexicon. (infra 4.1.1.) Thus, it will not be a problem for the prosodic modelling.

However, all the stresses will not have the same acoustic parameters. According to the linguistic context, the parameters will depend on many features: semantic, enunciative...

¹ Centre de Recherches Inter-langues sur la Signification en COntexte

2.2. Tunes

The tunes are global outlines that give to a phrase or to a typological sentence, a precise intonative pattern. They depend on the sentence modality (affirmative, negative, exclamative...) and on certain attitudes of the speaker (doubt, evidence, surprise...). We suggest a superpositional approach of intonation: the accents are considered as local prominences that are subordinated to tone units, sentences and paragraphs intonation.

2.3. Boundary tones.

Sense groups' limits can carry a specific accent. Bolinger (Bolinger, 1986) calls this phenomenon *boundary tone*. Even if the author denies any implication of the syntax in the prosody, we can notice that the sense groups coincide with syntactic groups. Syntax can thus influence intonative features. Furthermore, pauses' location is greatly determined by syntax.

2.4. Enunciative accents.

Enunciative accents highlight a specific word depending on its linguistic or its extralinguistic context. The numerous possibilities of their uses show us how complex the prediction of such accents are. However, these accents really contribute to the natural aspect of the synthetic speech.

While reading without monotony, the speaker should express emotive, attitudinal and enunciative factors. Expressions such as *nonsense*, *absurd* or *extremely* indicating a specific attitude of the speaker can have a particular accentuation. Locating these expressions are one of the main aspects of this work.

This classification of prosodic levels is a first approach to our problems. Now, we will detail the linguistic works that are necessary to a marked text including these different aspects.

3. METHODS FOR LOCATING ACCENTS

We can use several methods to predict accents location. We will study the different textual clues that we used for the marking of the texts: syntax, lexicon, argumental structure, and informational structure. The informational structure divides itself into two parts: the relative one and the absolute one. The absolute part deals with words which have a specific intonation by themselves: they do not need a particular context (ex: intensity markers). On the contrary, the relative part needs a specific context. It includes structures that place the reader in a state of expectation. It also includes incongruous utterances that can surprise the speaker.

3.1. Syntax

Syntax is probably the easiest way to analyse a given text automatically. Indeed, syntax, considered as a way to locate accents, can divide the text into two levels of syntactic structuring: phrases and syntagmas. An efficient synthesis needs such an analysis. Despite the problems of syntactic ambiguities, this method is efficient enough. Anyway, a disambiguation can easily be developed.

So, with this kind of analyses, it is possible to group words into chunks. It is then easier to predict the location of pauses and of potential boundary tones. (infra 4.2.3.).

3.2. Lexicon

To predict locations of accents, the lexicon is very important. We insert into it any information we need. (infra 4.1.1.). Besides grammatical and inflectional information, we also insert specific information. These information can locate words which carry an important semantic value. These words will generally carry a particular accent. For example, words like *tremendous*, *terrific*, *wonderful* are potentially carrying an emphatic accent.

Moreover, the lexicon helps us to locate neologisms. We work on the assumption that the more unexpected a word is, the more accented it will be. Thus, a word that is created by the author cannot be expected by the hearer or by the speaker (if the author is not the speaker himself). The intonation should underline this phenomenon.

Her sister was as unDursleyish as it was possible to be.

3.3. Argumental structure

We establish that the argumental structure of the sentence, that is to say the organisation between agents and a predicate, can act upon prosody. In close contact with informational structure, the category of subject and object can provide significant clues.

When the subject is expressed by a pronoun, the first complement or the first « fully semantic » verb often carries the main accent of the phrase.

In the following example, *improvement* and *become* will probably carry a more important accent due to their location in the phrases.

*Although this was an **improvement** in many ways it did **become** a bit depressing after a while...*

Evaluating the semantic value of subjects, objects and verbs, we can elaborate this prosodic level, once the text is disambiguated.

3.4. Informational structure

We have just concluded that argumental structure was a part of accentual prediction. This fact can be linked to informational structure of utterances. Indeed, the lexicon gives information that are necessary to an efficient understanding of an utterance. The modelling of the information is thus an important stake in prosody. We estimate that a new piece of information will be more prominent.

The expectation of the speaker is the main point. Incongruous or semantically strange sentences will then carry a specific accentuation. If a non-human noun is the subject of a verb which only accepts human subjects, it is logical to mark this incompatibility prosodically.

For example, in the following phrase: *A cat reading a map*. The noun *cat* or the verb *reading* will be accented more strongly because the situation cannot be foreseen.

The information given by lexicon and by contextual rules provide an overall comprehension of prosodic possibilities. To model and to test these possibilities, we applied these structures to the software Unitex and to the text-to-speech system Kali.

4. Implementation

The software Unitex provides us tools to locate linguistic structures efficiently. With the text-to-speech system Kali, we can listen to any text once acoustic parameters are merged into it.

4.1. Unitex

Unitex is a corpus processing system, based on automata-oriented technology. The concept of this software was born at LADL² at Paris VII, under the direction of its director, Maurice Gross. With this tool, you can handle electronic resources such as electronic dictionaries and grammars and apply them. You can work at three levels: morphology, the lexicon and syntax. Syntactic grammars can be represented with graphs and can help the user to make complex queries inside his corpus. Furthermore, we can merge some information into the text at some locations defined by these queries.

4.1.1. Lexicon

Dictionaries we used with Unitex are based on Electronic Dictionary DELAF (Silberztein, 93). The dictionary is made with the following formalism:

Lexical input	,	Lemma	.	Categories, phonologic or semantic information divided with +	:	Grammatical and inflection information
---------------	---	-------	---	---	---	--

Example:

was,be.V+Aux:3s

The main function of the DELA as we used it with Unitex is the insertion of semantic and syntactic information. This approach enables us to locate words depending on certain constraints. Launching the query <V-Aux>, we can then locate all the verbs which cannot be auxiliaries : *cut, read, inform*. But, it will not locate the following verbs: *be, have, will...*

For our purpose, we modified the DELA by adding semantic information, especially for emphatic words (*extremely, fantastic...*). When the word has several syllables, we also added the number of the syllable which usually carries the stress. (Ex: invention,.N+2:s).

4.1.2. Graphs

The graphs we can build on Unitex include labels which can be linked with one another. So, location of a specific context and insertion of markers in the text become possible. An Unitex graph can include subgraphs. The most interesting context for prosody can then be described by a large system of graphs.

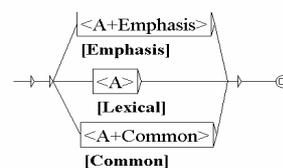
4.2. Systems created

Now, we present the systems of graphs which enable us to build marked texts for the study of prosody.

4.2.1. Word markers

Each syntactic category needs the building of a graph. The graph will merge into the text some specific markers before any occurrence of this category. In the following example, if the automaton that browses the text meets an adjective described in the lexicon as +Emphasis or +Common, the word markers [Emphasis] or [Common] will be merged. If the adjective is not described as an emphatic or as a common word, the mark [Lexical] will be merged.

Graph 1: Simplified graph for adjectives



² Laboratoire d'Automatique Documentaire et Linguistique

Lexicon 1: Extract of our English DELA.

wonderful,.A+Emphasis
 little,.ADV+A+Common
 french,.N+A

Text 1 will look like text 2 after the processing of lexicon 1 and graph 1

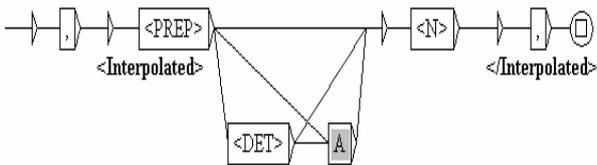
1: *It was a wonderful little French car.*

2: *It was a [Emphasis] wonderful [Common] little [Lexical] French car.*

4.2.2. Structure markers

The precise definition of the structures provided by the graphs is relevant and the software can then locate complex structures. For example, it is efficient to mark structure such as interpolated clauses.

Graph 2: Simplified graph to locate interpolated clauses.



We show here a simplified graph which can locate any prepositional syntagma between two commas. To mark this complex structure, we merge into the text an opening marker <Interpolated> and a closing one </Interpolated>. The sentence (1) will be marked as (2):

(1) : *Ron, at the next table, wasn't having much more luck*

(2) : *Ron,<Interpolated> at the next table</Interpolated>, wasn't having much more luck.*

4.2.3. Main graph

To take into account the argumental structure of the sentence, graphs representing a whole sentence or a whole phrase must be created. Through lack of time, we cannot explain all the graphs and subgraphs. However, they locate the prominent word which is related to argumental structure (supra 3.3). For these words, we use the markers <Main> ... </Main>.

4.3. Kali

4.3.1. The correspondences

We use two sorts of markers. The word markers are merged before each word to design whether they are lexical, emphatic, functional... The structure markers give the same influence to each word between the opening and the

closing markers. To establish the acoustic parameters, we created a table of correspondences where the parameters depend on the different markers. The parameters that are specific to Kali take into account : pitch, intensity and duration (Morel & al., 2001).

With a software, the markers are replaced by the acoustic parameters defined in the table. Once the text with the parameters is generated, we can test it with Kali.

4.3.2. Example

To illustrate this theory, we present an example of the processing applied to the following sentence: *He supposed this was some stupid new fashion.*

<Affirmative> [Functional] **He** <Main> [Lexical] **supposed** </Main> [Functional] **this** [Functional] **was** [Functional] **some** [Lexical] [Feeling] **stupid** [Lexical] [Common] **new** [Lexical] **fashion.** </Affirmative>

With the correspondences' table, we generate a text marked with acoustic parameters symbolised with + and - :

(-, -, -) **He** (++, ++, ++) **sup** (-, +, -) **posed** (-, -, -) **this**...

This generated text can then be voiced with the help of the software Kali.

5. CONCLUSION

With Unitex graphs and the correspondences, we can have a text with acoustic parameters merged before each syllable. Kali can interpret this presentation directly. We can thus replace a part of Kali processing by Unitex analyses. Furthermore, we can test some hypotheses concerning relations between semantics and prosody. We can also precisely formalize some enunciative data that would often remain uncertain without this kind of processing.

In the future, we intend to improve these analyses reinforcing the complementarities between oral and textual studies. Furthermore, some improvements, for the modelling of the informational structure are ought to be done. It will then be fundamental to compare written and spoken communication model.

REFERENCES

[1] D. Bolinger, *Intonation and its Parts: Melody in Spoken English*, 1986, Stanford University Press, Stanford, California.
 [2] M. Morel, A. Lacheret-Dujour, " Kali, synthèse vocale à partir du texte : de la conception à la mise en œuvre", *Traitement Automatique des langues*, 42 (1), pp.193-221, 2001.
 [3] M. Silberztein, *Dictionnaires électroniques et analyse automatique de textes : le système INTEX*, 1993, Masson, Paris, France.